

A Review on Outdoor and Indoor Automated Video Surveillance Systems

U. Pavan Kumar
Research Scholar
REVA University
Bengaluru

Bharathi S.H., PhD
Professor
REVA University
Bengaluru

ABSTRACT

Video surveillance is an important area of computer vision research, its applications including both outdoor and indoor automated surveillance systems. Detecting through video image processing is one of the most attractive alternative new technologies as it offers opportunities for performing substantially more complex tasks and providing more information than other sensors. Video Surveillance systems have as main goal to control the safety and the security of materials of which utilizing people. This paper provides an overview of various methods and techniques from the research area that address the problems of representation, recognition and learning of events, actions and activities of inhabitants from an environment.

Keywords

Video surveillance; tracking; Shadow removes; Motion detection.

1. INTRODUCTION

Video Surveillance is the most important in the field of security. The task of surveillance is to detect and track moving objects in the video sequence. Now-a-days, video surveillance system is an important security asset to control theft, trespassing or traffic monitoring, banks, department stores, highways, crowded public places, borders etc. Human Motion analysis helps in solving problems in indoor and outdoor surveillance applications. The research in the field of vision surveillance meets with the research in several techniques, especially in two important domestic aspects: home health, and home security. We identified two distinct directions where video surveillance techniques can be used in smart home environments: (a)ambient assisted living, whose goal is to improve the quality of life and to maintain independence of older and vulnerable people, and (b)vision-based fire detection, where video processing techniques are used to develop fire monitoring systems for smart home environments. In this paper, we focus attention on the ambient assisted living. An indoor surveillance system attempts to detect and recognize objects of interest from video obtained by one or more cameras along, eventually by fusion information obtained from cameras and other sensors installed in the monitored area. The aim of an automated video surveillance system is to understand what is happening in a monitored area and to take appropriate actions like alerting appropriate services if an undesirable occurred [2].

Ambient assisted living(AAL) can be described as information and communication technology based services and systems to provide older and vulnerable people with a secure environment, in order to improve their quality of life and reduce the costs of health and social care[3]. Video surveillance in smart home environments, attempts to detect, recognize and track persons from image sequences, and more importantly to understand and recognize their actions and activities. Video surveillance systems in smart home environments can be classified as: (a)centralized systems, where human activity analysis is performed on a centralized server, and (b)distributed systems, where analysis is performed on a distributed cameras network. In general, an in-house video surveillance system includes, as presented in Figure 1, the following modules: (a)motion detection,(b) object tracking, and (c)human motion analysis, in order to detect simple events, or recognize complex human(or group of humans) activities. The first two modules represent a syntactic low-level processing phase, while the last module represents a semantic high-level phase, which is able to detect and recognize human activities. The rest of this paper is organized as follows. Section 2 covers approaches related to low-level video processing techniques, such as motion detection and object recognition, while Section 3 covers current computer vision approaches for human motion analysis. In Section 4 explains selection of Region of Interest(ROI) , while Section 5 privacy issues in video surveillance systems are discussed. Finally conclusions are given in Section 6.

2. MOTION ESTIMATION AND OBJECT DETECTION

The low-level processing task of an indoor video surveillance system contains two distinct components, as presented in Figure 1: motion detection and object recognition. The input for these modules is a video sequence, each element from this sequence representing a video frame.

2.1 Motion Detection

Motion detection methods attempt to locate connected regions of pixels that represent the moving objects within the scene. Conventional approaches for motion detection methods use background subtraction, temporal differencing[4], and optical flow[5]. The first approach consists in subtraction of a background or reference model from the current image followed by a labeling process.

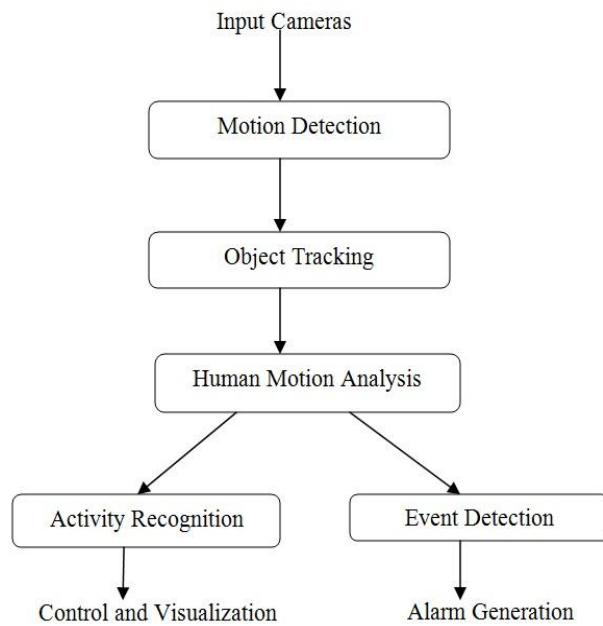


Fig 1: Flow of processing in a video surveillance system.

The second technique is based on the subtraction of two or more consecutive frames followed by thresholding. The optical flow approach uses the velocity field that warps one image into another image.

Background subtraction method is simple, but extremely sensitive to changes in dynamic scenes derived from lighting and extraneous events. Recent background subtraction algorithms focused on adapting to varying illumination conditions, geometry reconfiguration of background structure, and repetitive motion from clutter. Among others may include methods such as: (a) background subtraction methods for modelling a multiple modal background distribution[6], which use a Gaussian-based approach for real-time applications, (b) statistical background modelling[7], where an edge segment-based statistical background modelling is used, and (c) universal background subtraction algorithm for video sequences[8], which stores a set of values taken in the past in the same location or in the neighbourhood.

Temporal differencing is robust to dynamic changes of the environment, but generally this technique gives poor results in extracting all the relevant pixels, leaving holes inside moving objects. For this reason, background subtraction and temporal differencing techniques are used together in many approaches. Optical-flow-based methods can be used to detect independently moving objects even in the presence of camera motion, but the flow computation methods are computationally complex and very sensitive to noise, and cannot be applied to video streams in real time without using specialized hardware.

2.2 Object Tracking

The object tracking module is responsible for detecting and tracking of moving objects by using information from the motion detection module, object locations being subsequently transformed into 3D coordinates.

There are several approaches to classify object tracking methods. According to [9], object tracking can be classified into (i)region-based tracking, (ii)active contour-based tracking, (iii)feature-based tracking and (iv)model-based

tracking. We will use this classification because it is more appropriate with the in-house video surveillance methods.

Region-based tracking algorithms track objects according to variations of the image regions corresponding to the moving objects. In this case the background image is maintained dynamically, and motion regions are usually detected by subtracting the background from the current image. Feature-based tracking methods perform the tracking of objects by extracting first some elements from video frames, and then clustering them into higher level features, and finally by matching the features between frames.

Model-based tracking algorithms track objects by matching object models, produced with prior knowledge, to image data. The models are usually constructed off-line with manual measurement, or computer vision techniques.

We focus on model-based human body tracking because this method is related to in-house vision surveillance systems. Model-based human body tracking is known as analysis-by-synthesis, and it is used in a predict-match-update style. In a first step, the pose of the model for the next frame is predicted according to prior knowledge and tracking history, then the predicted model is synthesized and projected into the image plane for comparison with the image data.

Model-based human body tracking techniques are based on three components: (i)human body models, (ii)motion models, and (iii)search strategies.

Human body models can be represented in several styles: (a) a stick figure representation, which can be used to build a hierarchical model of human dynamics encoded using Hidden Markov models[10], (b)2D contour, which represents the human body projections in an image plane (the human body segments can be modelled by 2D ribbons[11], or blobs[12], (c)volumetric models that use 3D models such as elliptical cylinders, cones, spheres, super-quadratics, etc[13], and (d) hierarchical model[14] that include several levels (skeleton, ellipsoid meatballs, polygonal surface) for achieving more accurate results.

Motion models of human limbs and joints are widely used in tracking, because the movements of the limbs are strongly constrained. These motion models serve as prior knowledge to predict motion parameters[15], to interpret and recognize human behaviours[16], or to constrain the estimation of low-level image measurements[17].

Search strategies are used to reduce the solution space, because pose estimation in a high-dimensional body configuration space is intrinsically difficult. Generally, there are four main classes of search strategies: dynamics, Taylor models, Kalman filtering, and stochastic sampling. Dynamical strategies use physical forces applied to each rigid part of the 3D model of the tracked object that guide the minimization of the difference between the pose of the 3D model and the pose of the real object[13]. Taylor models use an incremental approach to improve an existing estimation, using differentials of motion parameters with respect to the observation to predict better search directions[18]. The Kalman filter is a recursive linear estimator used for tracking the shape and position over time in which the density of the motion parameters can be modelled as Gaussian[19]. Stochastic sampling strategies, such as Markov Chain Monte Carlo[20], and condensation[21], are used to handle clutter that causes the probability density function for motion parameters to be multimodal and non-Gaussian.

Compared with other tracking algorithms, model-based tracking algorithms have several advantages such as:

These algorithms are robust (by using the prior knowledge of the 3-D contours or surfaces of objects), and they can obtain better results even under occlusion.

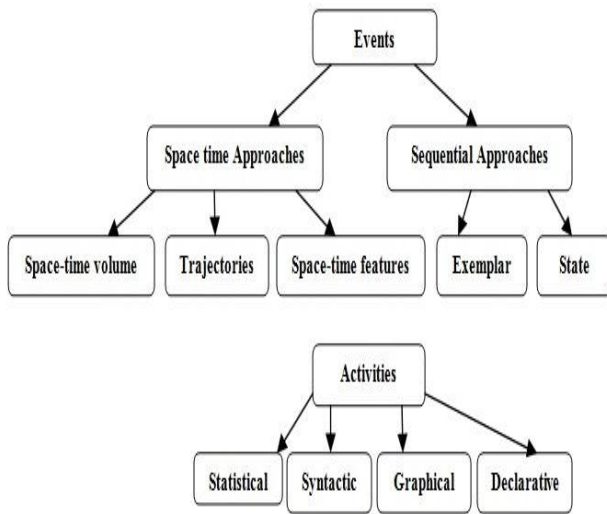


Fig 2: Overview of approaches for event and recognition

The structure of human body, the constraint of human motion, and other prior knowledge can be fused in a single information structure.

The 3-D model-based tracking algorithms can be applied even when objects greatly change their orientations during the motion.

However model-based tracking algorithms have two main disadvantages: the necessity of constructing the models, and the high computational cost.

3. HUMAN MOTION ANALYSIS

After the of motion detection step, human actions can be viewed as a series of detected motions. In[22] an action taxonomy is defined, based on the degree of abstraction:

- (a) basic motion recognition, or action primitives, which represent the atomic entities out of which actions are built,
- (b) a set of different or repetitive action primitives that make up an action, and
- (c) activities, which represent complex sequence of actions performed by several humans who could be interacting with each other in a constrained manner. In the following we use the term of event instead of action. Various approaches for categorizing recognition methods for events and activities have been proposed[2], [23], [24], based on the approaches used for event modelling.

Figure 2 presents an hierarchy that is similar to the approach proposed in[24], where the notion of single-layered approach is replaced with the notion of event, and hierarchical approach is replaced with the notion of activity. Single-layered approaches from[24] represent methods that recognize human motion directly based on sequences of images, being suitable for recognition of events, while hierarchical approaches represent high-level human activities that can be described in term of sub-events.

3.1 Modeling and Recognition of Events

Methods for event recognition can be again classified into two types, as presented in Figure 2, depending on the model used to describe human motion: (i) space-time approaches, where the input video is considered as a 3D volume, and

(ii) sequential approaches, where the input is considered as a sequence of frames.

3.1.1 Space Time Approaches

The aim of these approaches is to recognize human activities by analysing space-time volumes from the video frames. They are divided into several categories based on the type of features considered in the 3D space-time volumes:

- (a) volumes, (b) trajectories, (c) local features

Volumes are viewed themselves as features, and the recognition is performed by measuring the similarity between two volumes. In[25] a template matching method is applied to a pair of (MEI, MHI), where MEI represents the motion-energy of an image, and MHI represents the motion-history of the image. This system is able to recognize simple actions like sitting, arm waving, and crouching. Authors of[27] introduce a method that uses background modeling and spatio-temporal template matching based technique. The motion history images are used to construct object shape information for recognizing different human activities, such as walking, standing, bending, sleeping and jumping. The space-time volume approaches have difficulties in recognizing actions when multiple persons are present in the scene, or when the actions cannot be spatially segmented. Trajectories are methods, where features are represented as a sequence of 2D or 3D points corresponding to positions of the moving objects, and human body part estimation methodologies are then used to extract the joint positions of a person at each image frame. One major advantage of the trajectory-based approaches is the ability to analyse detailed levels of human movements. The method measures structural similarity between two videos, which enables the detection and localization of complex-structured activities.

3.1.2 Sequential Approaches

These approaches consider as input a sequence of video frames, and determine if an activity has occurred in the video when a particular sequence of features characterizing the activity is observed. Depending on the method used to recognize human motion, they are classified into (i) exemplar-based, and (ii) model-based.

Exemplar-based approaches describe classes of human actions using training samples, and represent human activities by maintaining a template sequence or a set of sample sequences of actions. The dynamic time warping (DTW) algorithm has been adopted for matching two sequences with variations[35]. In[36] human activities are recognized by modeling them as linear time invariant (LTI) systems. The system converts a sequence of images into a sequence of silhouettes, extracts contour information, and then uses SVMs for classifying a new video input. In this way some simple actions such as 'slow walk' and 'fast walk' can be recognized. In[37] a method for the analysis of human motion has been proposed that uses joint estimation of human behavioural information, such as head and body pose.

In model-based approaches the human motion is represented as a model composed of a set of states. The model is statistically trained, and one statistical model is constructed for each activity. Hidden Markov models(HMMs) and

dynamic Bayesian networks(DBNs) have been widely used for state model-based approaches. In[38] a DBN is used to recognize gestures of two interacting persons. Some gestures such as 'stretching an arm' and 'turning a head left' can be recognized by constructing a tree-structured DBN. In[39] an efficient recognition algorithm using coupled hidden semi-Markov models(CHSMs) has been proposed, which extends the coupled HMMs by modeling the duration of an activity staying in each state. In[40] a method for object detection and action recognition is presented that uses a Histogram of Oriented Gradients(HOG) method for object detection, and a Hidden Markov Model(HMM) for capturing the temporal structure of the features. Exemplar-based approaches provide more flexibility for recognition of human activities than the model-based approaches(e.g. such systems can maintain multiple and different sample sequences). In addition, the amount of necessary training data can be less than in the case of state model-based approaches. On the other hand, state-based approaches are able to perform probabilistic inference for activity determination.

3.2 Modeling and Recognition of Activities

Higher-level representation and reasoning methods are necessary for modeling complex activities having inherent structure and semantics. Accordingly to this assumption, in approaches for human activity recognition (or hierarchical approaches[24]) a high-level activity is considered as composed of several simpler activities, denoted by sub-activities (or actions, or events).

As presented in Figure 2, approaches for modeling and recognition of activities are divided into four categories:

(i)statistical, (ii)graphical based, (iii)syntactic, (iv)knowledge based.

3.2.1 Statistical Approaches

In the case of statistical approaches, multiple layers of state-based models such as HMMs are used to recognize activities with sequential structures. At the bottom layer, atomic actions are recognized from sequences of feature vectors, while the second-level treats this sequence of atomic actions as observations generated by the first-level. For each model, a classifier is constructed by using the maximum likelihood estimation(MLE), or the maximum a posteriori probability (MAP). The method of multi-layered HMMs has been explored by various researchers. In[41] hierarchical HMMs of two layers are constructed to recognize human activities such as 'a person having a meal', or 'a person having a snack'. In [42] a multi-layered HMM is constructed to recognize group activities occurring in a meeting room. The system recognized atomic actions of 'speaking', 'writing', and group activities such as 'monologue', 'discussion', and 'presentation'.

3.2.2 Graphical based Approaches

These approaches contain a graphical representation of some state-based models. Among others we can mention belief networks and Petri nets. Belief networks represent Bayesian networks that specify complex conditional dependencies between a set of random variables that are encoded as local conditional probability densities. In addition, Dynamic Belief networks(DBNs) represent a generalization of the Bayesian networks by incorporating temporal dependencies between random variables, and approaches using DBNs may contain multiple levels of hidden states. In[43] DBNs are constructed to recognize group activities in a conference room, such as 'break', 'presentation', and 'discussion', based on atomic actions such as 'talking', or 'asking'. In[44] a method for

interpretation of human-object interactions is presented, that integrates information from perceptual tasks such as human motion analysis, and object detection.

3.2.3 Syntactic Approaches

These approaches are based on grammars that express the structure of a process using a set of production rules. In video motion understanding approaches, terminal symbols correspond to abstraction primitives, non terminal symbols correspond to semantic concepts and production rules correspond to the semantic structure of the activities. The recognition of an activity represents a parsing process. There are several approaches for using grammars in video activity understanding, which use different type of grammars: context-free grammars, stochastic grammars, and attribute grammars.

Context-free grammars(CFG) are one of the earliest used type of grammars for video activity recognition[48]. In[49] a method that use the CFG formalism to model and recognize composite human activities and multi-person interactions is proposed. This method represents a hierarchical approach, having the lower-levels composed of HMMs and BNs, while the higher-level interactions are modelled by CFGs. In complex scenarios involving several persons, and requiring temporal relations, it is difficult to formulate the grammatical rules manually. Learning the rules of the grammar from training data is a promising alternative, but it has proved to be extremely difficult in the general case [50].

Stochastic grammars allow probabilities to be associated with each production rule. This extension provides a mechanism to deal with the uncertainty inherent in video surveillance systems. In[51] Stochastic context-free grammars (SCFG) were used to model the semantics of activities whose structure was assumed to be known, where low-level primitive detection is performed by using HMMs. In[52] SCFGs are used to model multi-tasked activities (i.e., activities that have several independent threads of execution with intermittent dependent interactions). As in the case of CFGs, approaches for learning the structure of SCFGs have been proposed. In[53] a method for automatically learning a Probabilistic Context-free grammar for recognizing human actions from sequences of human silhouettes is presented. Attribute grammars associate conditions with production rules. In video-surveillance systems, attributes can be attached to primitive events. In[54] Probabilistic Attribute Grammars have been used for multi-agent activities in video surveillance. The primitive events used in this approach are 'appear', 'disappear', 'stop', and 'start'. The primitive events are associated with attributes such as id(identity of the entity involved), loc(location where the primitive event occurs), and class(object classification label).

3.3 Knowledge based Approaches

Knowledge-based approaches represent activities in terms of primitives and constraints on them. These methods can express more complex constraints than grammar-based approaches.

Knowledge approaches can be divided in two categories: logic-based approaches, and ontology-based approaches. Logic-based approaches use logical rules to describe common-sense domain knowledge that is used to describe human activities. In[55] a method for representing activities using declarative models is presented, recognition being performed by using a constraint-satisfaction algorithm. In[56] a hierarchical representation to recognize actions performed by a single person is presented. The first level realizes

detection and tracking of moving objects from the video stream, while the second level takes the inferred trajectory of each object, and the contextual information, to recognize the behaviour of the moving objects among a large number of potential scenarios. In [57] this representation is extended by considering an activity to be composed of several action threads, each action thread being modelled as a stochastic finite-state automaton.

The main advantage of the statistical and syntactic approaches is the fact that they allow reliable recognition with noisy inputs, while the main drawback is the difficulty for representing and recognizing activities with concurrently organized sub-events. The main advantage of the graphical-based and declarative approaches is the possibility to represent and recognize human activities with complex temporal structures, while the major drawback is the deterministic structure at the high level component.

4. SELECTION METHOD OF ROI

Selection methods of ROI which can be classified into one of the following three categories:

1) Frames subtraction method; 2) Background update method; 3) Virtual loop method.

The objective of select methods of ROI is to find possible location in an image quickly further detection and tracking.

4.1 Frames Subtraction Method

Frames subtraction methods employ a threshold technique over the inter-frame difference, where pixel differences has considered. These techniques utilize the double difference on two or more successive frames to detect a ROI. The advantage of these methods is that it's an adaptive method. The major problem of these techniques is that it depends too much on the time interval of selection the successive frames. Here below mentioned some representative approaches including inter-frame difference method, statistical test method and spatial Markov random field methods.

4.1.1 Inter-frame Difference Method

Inter-frame difference method has been used frequently for selection of ROI in computer vision [35][36][37]. Frames subtraction methods detect moving objects region from current input image by performing a difference on two or three consecutive inter-frames. A criterion is provided to decide the ROI. The difference map is usually binarized using a predefined threshold value to obtain the motion region. It is easy to realize and its low computation. The advantage of these methods is low efficiency, especially where there is no prior knowledge. To overcome the shortcoming of these methods, adapted threshold techniques are proposed [37].

4.1.2 Statistical Test Method

It is pixel-wise independent decision, as a step forward of Inter-frame difference methods [38]. The method which assume intrinsically that the detection of temporal changes is equivalent to the motion detection. The shortcoming of this method is that this assumption is valid when either large displacements appear are sufficiently textured. To overcome this drawback, temporal change detection masks and filters have also been introduced [36]. The application of these masks improves the efficiency of the change detection algorithms, especially in the case where a prior knowledge about the size of the moving objects is available.

4.1.3 Spatial Markov Random Field Method

In order to avoid the limitation of prior knowledge about the size of the object, spatial Markov random field method is provided [40]. Motion detection is considered as a statistical estimation problem in this method. It can be used to smoothness of image regions.

4.2 Background Update Method

Background update techniques are mostly used for detection of motion in many real-time detection and [41][42][43] tracking applications, including frame average method, selective updating method, minimum and maximum intensity value method, mixture of Gaussian method and k-means clustering techniques. In these approaches, difference between the future frame and the constructed background image is performed to detected foreground objects. To achieve accurate results over time the update process must be able to adapt to lighting changes and to handle new objects that appear in a scene.

4.2.1 Frame Average Method

Frame average method is a technique for detection [44]. The main point of it is to establish background frame in order to separate the ROI from current image and then ROI can be detected. Background is established by calculating average value of several consecutive frames in these methods. The main drawback of this method is that it's difficult to detect stationary objects. The key point of these methods is that how to update the background adaptively.

4.2.2 Selective Updating Method

Only selective region of frames are updated for selective updating method. Wang [45] proposed a method that models the background into dynamic regions and a quasi-static region. A block-based Eigen-space approach is used to model the quasi-static background. The advantage of these methods is that it improves the robustness. However, this technique needs prior knowledge to confirm the region to update.

4.2.3 Minimum and Maximum Intensity Value Method

In these methods [39], minimum and maximum intensity values, and maximum temporal derivative for each pixel are stored to initialize background model. The background model is periodically updated by using a combination of pixel-based method and object-based method. Kalman and Brandt [37] propose an adaptive background model using Kalman Filtering to adapt to the temporal variation of weather and environmental illumination. Stauffer and Grimson [38] used a mixture of normal distributions to model a multi-model background image sequence.

4.2.4 Mixture of Gaussian Method

The pixel-level mixture of Gaussians (MOG) background model has become very popular because of its efficiency in modeling multi-modal distribution of background (such as waving trees, light reflection, etc.). It can adapt to a change of the background and implement the method in real time potentially. The basic idea is to assume that the time series of observations, at a given pixel, is independent of the observations at other image pixels [50]. Friedman and Russell [51] modeled the intensity values of a pixel by using a mixture of three Normal Distributions and applied the proposed method to traffic surveillances. Stauffer and Grimson [44] presented a method that models the pixel intensity by a mixture of K Gaussian distributions.

4.2.5 K-means Clustering Techniques

Cluster analysis can be performed on the co-efficient space to build a self-consistent aggregation of many individual. By taking into account changing pixels of vehicle region, vehicle geometry can be estimate from stable video sequences. This method doesn't need a prior calibration of the image sequences.

4.3 Virtual Loop Methods

Virtual Loop methods exploit the concept of inductive loop to detect vehicle passing by monitoring illumination change in pre-specified region of a frame during detection processing. It employ prior knowledge to select the ROI directly[46]. The Selection of ROI is based on the decision of human based on the need of detection requirement. In general, lanes are always selected as ROI. As the kind of processing checks the pre-specified regions of frame only, its processing speed is fast.

5. PRIVACY AND VIDEO SURVEILLANCE SYSTEM

The variety of the approaches for video surveillance in smart home environments, and of advances in camera and computing equipment hardware in recent years, there is a real barrier to introduce the video surveillance system at home. The most relevant problem is the total lack of privacy. Actually relatively few surveillance systems addressing privacy issues have been published.

The majority of work on image sequences applies simple distortion methods such as pixelation or blurring to obfuscate parts or all of the image[53]. The Privacy Cam architecture proposed in[54] suppresses automatically segmented foreground objects in the scene and cryptographically secures access to the altered video stream produced by the system. In [55] it is proposed a method for face de-identification by using the k-Same algorithm, which offers formal privacy protection guarantees. In[56] an interesting approach is proposed, the respectful camera, where the problem of automatically obscuring faces in real time to assist in video privacy enforcement is considered.

This approach targets surveillance applications where an alarm need to be detected, and persons are required to wear colored hats or vests. Authors from[57] present a privacy preserving surveillance video system where the privacy information can be stored and retrieved in a secure and reliable manner. A digital signature is inserted into the resulting video for authentication.

6. CONCLUSIONS

In this paper, presented an overview of the current approaches on vision surveillance techniques. Depending on the range of interest, different methods are proposed to solve fixed problems. The objects are removed rapidly by direct updating of model while preventing abandoned objects from decomposing into background. In this review, summarized the methodologies used in several stages of indoor and outdoor video surveillance system: motion detection, object tracking, and human activity recognition. Further working on improving the appearance of moving as well as non moving objects of proposed algorithms.

7. REFERENCES

[1] Ricquebourg, D.Menga, D.Durand, B.Marhic, L. Delahoche, and C.Loge, "The smart home concept: our immediate future," in Proc.IEEE Int.Conf.on E-Learning in Industrial Electronics, dec2006, pp. 23–28.

[2] G.Lavee, E.Rivlin, and M.Rudzsky, "Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video," *Trans.Sys. Man Cyber Part C*, vol.39, no.5, pp.489-504, sep 2009.

[3] F.Cardimax, D.Bhowmik, C.Abhayaratne, and M.S.Hawley, "Video based technology for ambient assisted living: A review of the literature," *Journal of Ambient Intelligence and Smart Environments*, vol.3, no. 3, pp.253-269, 2011.

[4] A.J.Lipton, H.Fujiyoshi, and R.S.patil, "Moving target classification and tracking from real-time video," in Proc. IEEE Workshop Applications of Computer Vision, 1998, pp.8-14.

[5] J.Barron, D.Fleet, and S.Beauchemin, "Performance of optical techniques," *Intl. Journal of Computer Vision*, vol.12, no.1,pp.42-77, 1994.

[6] M.Hedayati, W.Zaki, and A.Hussain, "Real-time background subtraction for video surveillance: From research to reality," in Proc. Int. Colloquium on Signal Processing and Its Applications, 2010, pp. 1–6.

[7] M.Murshed, A.Ramirez, and O.Chae, "Statistical background modeling: An edge segment based moving object detection approach," in Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, 2010, pp. 300–306.

[8] O.Barnich and M.V.Droogenbroeck, "Vibe":A universal background subtraction algorithm for video sequences," *IEEE Trans. on Image Processing*, vol.20, no.6, pp.1709-1724, 2011.

[9] W.Hu, T.Tan, L.Wang, and S.Maybank, "A survey on visual surveillance of object motion and behaviors", *IEEE Trans. on Systems, Man, and Cybernetics, Part C*, vol. 34, no. 3, pp. 334–352, 2004.

[10] I.A.Karaulova, P.M.Hall, and A.D.Marshall, "A hierarchical model of dynamics for tracking people with a single video camera," in Proc. British Machine Vision Conf., 2000, pp. 261–352.

[11] S.A.Niyogi and E.H.Adelson, "Analyzing and recognizing walking figures in XYT," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, 1994, pp. 496–474.

[12] Q.Delamarre and O.Fagugeras, "3D articulated models and multi-view tracking with physical forces," *Comput. Vis. Image Understanding*, vol.81, no.3, pp.328-357, 2001.

[13] R.Plankers and P.Fua, "Articulated soft objects for video based body modeling," in Proc. Int. Conf. Computer Vision, Vancouver, BC, Canada, 2001, pp. 394–401.

[14] T.Zhao, T. S. Wang, and H. Y. Shum, "Learning a highly structured motion model for 3D human tracking," in Proc. Asian Conf. Computer Vision, Melbourne, Australia, 2002, pp. 144–149.

[15] C.Bregler, "Learning and recognizing human dynamics in video sequences," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997, pp. 568–576.

[16] H.Sidenbladh and M.Black, "tochastic tracking of 3D human figures using 2D image motion," in Proc.

- European Conf. Computer Vision, Dublin, Ireland, 2000, pp. 702–716.
- [17] D.G.Lowe, "Fitting parameterized 3-D models to images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 441–450, 1991.
- [18] J.Hoshino, H.Saito, and M.Yamamoto, "A match moving technique for merging CG cloth and human video," *J.Visualiz. Comput. Animation*, vol.12, no.1, pp.23–29, 2001.
- [19] J.E.Bennett, A.Racine-Poon, and J.C.Wakefield, "Mcmc for nonlinear hierarchical models," in *Markov Chain Monte Carlo in Practice*, W.R.Gilks, S.Richardson, and D.J. Spiegelhalter, Eds. U.K:Chapman and Hall, 1996, pp. 339–357.
- [20] M.Isard and A.Blake, "CONDENSATION conditional density propagation for visual tracking," *Int J.Comput Vis.*, vol.29, no.1, pp.5-28,1998.
- [21] T. B. Moeslund, A. Hilton, and V.Krger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Understanding*, vol. 104, pp. 96–126, 2006.
- [22] P.Turaga, R.Chellappa, V.Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE on Circuits and Systems for Video Technology*, vol. 18, pp. 1473–1488, 2008.
- [23] J.Aggarwal and M.Ryoo, "Human activity analysis: A review machine recognition of human activities: A survey," *ACM Computing Surveys*, vol. 13, no. 3, 2011.
- [24] A.Bobick and J.Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 3, pp. 257–267, 2001.
- [25] Y.Ke, R.Sukthankar, and M.Hebert, "Spatio-temporal shape and correlation for action recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [26] C.M.Sharma, A.K.S.Kushwaha, S.Nigam, and A. Khar, "Automatic human activity recognition in video using background modeling and spatio-temporal template matching based technique," in *Proc. of the Int. Conf. on Advances in Computing and Artificial Intelligence*, 2011, pp. 97–101.
- [27] Y.Sheikh, M.Sheikh, and M. Shah, "Exploring the space of a human action," in *Proc. IEEE Int. Conf. on Comput. Vision*, vol. 1, 2005, pp. 144–149.
- [28] C.Rao and M.Shah, "View-invariance in action recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, 2001, pp. 316–322.
- [29] M.Blank, L.Gorelick, E.Shechtman, M.Irani, R., and Basri, "Actions as space-time shapes," in *Proc. IEEE Int. Conf. on Comput. Vision*, vol. 2, 2005, pp. 1395–1402.
- [30] J.C.Niebles, H.Wang, and L.Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *Int. J. Comp. Vis.*, vol. 79, no. 3, pp. 299–318, 2009.
- [31] L.Zelnik-Manor and M. Irani, "Event-based analysis of video," in *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*, vol. 2, 2001, pp. 123–130.
- [32] A.Wiliem, V.Madasu, W. Boles, and P.Yarlagadda, "Context space model for detecting anomalous behaviour in video surveillance," in *Proc. IEEE Int. Conf. on Information Technology. Generations*, 2012, pp. 18–24.
- [33] A.Veeraraghavan, R.Chellappa, and A.Roy-Chowdhury, "The function space of an activity," in *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*, vol. 1, 2006, pp. 959–968.
- [34] R.Lublinerman, N.Ozay, D.Zarपालas, and O. Camps, "Parameterized modeling and recognition of activities," in *Proc.Int. Conf. on Pattern Recognition*, 2006, pp. 347–350.
- [35] N. Diehl, *Object-Oriented Motion Estimation and Segmentation in Image Sequences*, *IEEE Trans. Image Processing*, vol. 3, pp. 1,9011,904, Feb. 1990.
- [36] Nikos Paragios, Rachid Deriche *Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22 No.3 pp 266-280, March, 2000.
- [37] Lei Xie; Guangxi Zhu; Yuqi Wang; Haixiang Xu; Zhenming Zhang; *Robust vehicles extraction in a video-based intelligent transportation systems* *IEEE 2005 International Conference on Communications, Circuits and Systems*, Volume 2, 27-30 May 2005 Page.887-890.
- [38] H.H.Nagel, G.Socher, H.Kollnig, and M.Otte, *Motion Boundary Detection in Image Sequences by Local Stochastic Tests*, *Proc. European Conf. Computer Vision*, vol. II, pp. 305-315, 1994.
- [39] T.Aach and A.Kaup, *Bayesian Algorithms for Adaptive Change Detection in Image Sequences Using Markov Random Fields*, *Signal Processing: Image Comm.*, vol. 7, pp. 147-160, 1995.
- [40] J.M.Odobeze and P.Bouthemy, *Robust Multi resolution Estimation of Parametric Motion Models*, *J.Visual Comm. and Image Representation*, vol. 6, pp. 348-365, 1995.
- [41] N. Paragios and G.Tziritas, *Adaptive Detection and Localization of Moving Objects in Image Sequences*, *Signal Processing: Image Comm.*, vol. 14, pp. 277-296, 1999.
- [42] Kalman.K. and BrandtA., *Moving Object Recognition Using an Adaptive Background*, In *Proc. Time-varying Image Processing and Moving Object Recognition*, Amsterdam, Netherlands, pp 289-296, 1990.
- [43] Stauffer C. and Grimson W., *Adaptive Background Mixture Models for Real-time Tracking*. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, 246-252, 1999.
- [44] Baisheng Chen, Yunqi, Lei, and Wangwei Li, *A Novel Background Model for Real-time Vehicle Detection*. *IEEE ICSP'04*, 12761279, 2004.
- [45] Zhang Jinglei; Liu Zhengguang; *A Vision-Based Road Surveillance System Using Improved Background Subtraction and Region Growing Approach* *IEEE Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing* Volume 3, July 30 2007-Aug. 1 2007 Page(s): 819 – 822

- [46] C.C.C.Pang,, W. W. L. Lam, and N. H. C. Yung. "A Novel Method for Resolving Vehicle Occlusion in a Monocular Traffic-image Sequence," IEEE Trans. on Intelligent Transportation Systems, vol. 5, no. 3, pp.129-141, Sep. 2004.
- [47] A.Mittal and N.Paragios."Motion-Based Background Subtraction using Adaptive Kernel Density Estimation," CVPR,, Washington, DC, 302-309, 2004.
- [48] Junxian Wang, George Bebis and Ronald Miller, Overtaking Vehicle Detection Using Dynamic and Quasi-Static Background Modeling, IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005,
- [49] Haritaogu I., Harwood D. and Davis L., W. Real-Time Surveillance of People and Their Activities, IEEE Transaction on Pattern Analysis and Machine Intelligence. 22, (8), 809-830, August, 2000.
- [50] N.Friedman and S.Russell. "Image Segmentation in Video Sequences. A Probabilistic approach, "Thirteenth Conf. on Uncertainty in Artificial Intelligence, pp. 175-181,1997.
- [51] Hanzi Wang; Suter, D.; A re-evaluation of mixture of Gaussian background modeling ICASSP '05). IEEE International Conference on video signal processing applications Vol. 2 Page(s) 1017 - 1020.
- [52] Andrew H.S. Lai and Nelson H.C.Yung, "Vehicle-type Identification Through automated virtual loop assignment and block-based direction biased motion estimation", IEEE Transactions on Intelligent Transportation Systems, vol.1,no.2, pp.86-97, Jun. 2000.
- [53] C.Neustaedter, S. Greenberg, and M. Boyle, "Blur filtration fails to preserve privacy for home-based video conferencing," ACM Trans. on Computer-Human Interaction, vol. 13, no. 1, pp. 1–36, 2005.
- [54] A.Senior, S.Pankati, A.Hampapur, L.Brown, Y. Tian, and A.Ekin, "Blinkering surveillance: enabling video surveillance privacy through computer vision," IEEE Security and Privacy, vol. 3, no. 5, pp. 50–57, 2005.
- [55] E. Newton, L.Sweeney, and B.Malin, "Preserving privacy by de-identifying facial images," IEEE Trans. on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232–243, 2005.
- [56] J.Schiff, M.Meingast, D.Mulligan, S.Sastry, and K.Gold-berg, "Respectful cameras: Detecting visualmarkers in real time to address privacy concerns," in Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007,pp. 971–978.
- [57] W.Zhang, S.Ching S.Cheug, and M. Chen, "Hiding privacy information in video surveillance system," in Proc. IEEE Intl. Conf. on Image Processing, 2005, pp. 868-871.